

Dynamics of Normalization Underlying Masking in Human Visual Cortex

Jeffrey J. Tsai,^{1,2} Alex R. Wade,^{1,2,3} and Anthony M. Norcia⁴

¹Smith-Kettlewell Eye Research Institute, San Francisco, California 94115, ²Department of Neurology, University of California San Francisco, San Francisco, CA 94143, ³Department of Psychology, University of York, Heslington, YO10 5DD, United Kingdom, and ⁴Department of Psychology, Stanford University, Stanford, California 94305

Stimulus visibility can be reduced by other stimuli that overlap the same region of visual space, a process known as masking. Here we studied the neural mechanisms of masking in humans using source-imaged steady state visual evoked potentials and frequency-domain analysis over a wide range of relative stimulus strengths of test and mask stimuli. Test and mask stimuli were tagged with distinct temporal frequencies and we quantified spectral response components associated with the individual stimuli (self terms) and responses due to interaction between stimuli (intermodulation terms). In early visual cortex, masking alters the self terms in a manner consistent with a reduction of input contrast. We also identify a novel signature of masking: a robust intermodulation term that peaks when the test and mask stimuli have equal contrast and disappears when they are widely different. We fit all of our data simultaneously with family of a divisive gain control models that differed only in their dynamics. Models with either very short or very long temporal integration constants for the gain pool performed worse than a model with an integration time of ~ 30 ms. Finally, the absolute magnitudes of the response were controlled by the ratio of the stimulus contrasts, not their absolute values. This contrast–contrast invariance suggests that many neurons in early visual cortex code relative rather than absolute contrast. Together, these results provide a more complete description of masking within the normalization framework of contrast gain control and suggest that contrast normalization accomplishes multiple functional goals.

Introduction

The stereotyped cytoarchitecture of the neocortex suggests that similar neural circuitry, and therefore similar computations, might be found across different areas of the brain (Creutzfeldt, 1977; Douglas and Martin, 2004). Divisive normalization is one such computation in which excitatory inputs to a cell population are modeled by polynomial terms that form the numerator of the computational operator (Anderson et al., 2000; Miller and Troyer, 2002; Kouh and Poggio, 2008). These responses are divided, or “normalized”, by inhibitory inputs (Riesenhuber and Poggio, 1999). At a descriptive level, normalization is successful in explaining numerous visual perceptual phenomena, including contrast adaptation (Greenlee and Heitger, 1988; Heeger, 1992), pattern masking (Foley, 1994; Candy et al., 2001), attentional modulation (Boynton, 2009; Reynolds and Heeger, 2009)—all of which may be viewed as manifestations of gain control. Gain control is an essential mechanism for adjusting a system’s sensitivity for efficient (Schwartz and Simoncelli, 2001) and robust

(Carandini, 2007; Carandini and Heeger, 2011) representation of the external world.

An important perceptual phenomenon in which gain control plays a central role is masking. In masking, the detectability of a stimulus is reduced by other stimuli presented to the same or similar region of visual space (Legge and Foley, 1980). Masking is prevalent because objects in the natural environment are inevitably observed in context rather than in isolation. Neural correlates of masking have been observed in single cells in visual cortex, where the response to a preferred stimulus is reduced by the superimposition of a second stimulus that by itself elicits little or no response (Morrone et al., 1982; Bonds, 1989; DeAngelis et al., 1992; Carandini, 2004). Masking, or “suppression,” can be so prominent that the neural activity generated by a stronger stimulus can completely dominate that of a weaker stimulus. This behavior is a neural correlate of a winner-take-all (WTA) computational operator (Kouh and Poggio, 2008). Busse et al. (2009) showed that WTA behavior could be observed at a neural population level and modeled by a divisive normalization process.

Here, we characterize masking using a frequency-domain nonlinear analysis method and a neural population correlate of masking obtained from source-imaged EEG. Two frequency-tagged stimuli generate response components at frequencies that are low-order sums and differences of the input frequencies. The specific frequency components that are observed and their relative amplitudes depend very strongly on the underlying nonlinearity and comprise a “fingerprint” of this computation (Victor and Shapley, 1980; Regan and Regan, 1988). Regan and Regan

Received Sept. 1, 2011; revised Jan. 2, 2012; accepted Jan. 6, 2012.

Author contributions: J.J.T. and A.M.N. designed research; J.J.T. performed research; J.J.T. analyzed data; J.J.T., A.R.W., and A.M.N. wrote the paper.

This research was supported by the National Eye Institute of NIH (K23EY020876 to J.J.T. and R01EY018157 to A.W.). We thank M. Carandini, D. Mannon, and P. Verghese for comments on the manuscript.

The authors declare no financial conflicts of interest.

Correspondence should be addressed to Jeffrey J. Tsai, Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115. E-mail: jeff@ski.org.

DOI:10.1523/JNEUROSCI.4485-11.2012

Copyright © 2012 the authors 0270-6474/12/322783-07\$15.00/0

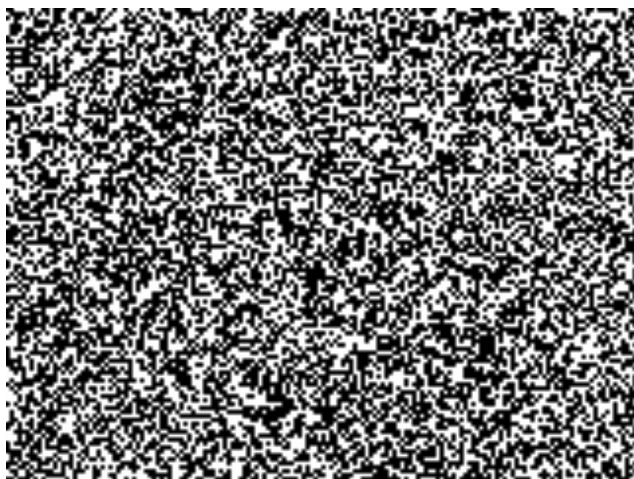


Figure 1. Random noise pattern used as visual stimulus. The pattern has a mean intensity of 0.5 (1 being white and 0 being black). Each pattern is multiplied by a temporal sinusoid $\frac{1}{2}c_i \sin(2\pi f_i t)$, where c_i and f_i are the contrast and frequency, respectively, and added to a uniform background of mean intensity of 0.5. Two patterns are superimposed by displaying interleaved video lines.

(1988) explored the effects of different forms of static nonlinearity, including a sigmoidal nonlinearity, on the sum of two sinusoids. They showed analytically that this resulted in equal second-order difference and sum responses. Candy et al. (2001) used a frequency tagging approach to study both iso-orientation and cross-orientation masking in human visual evoked potentials (VEP). They studied a limited range of input contrasts and found that trends in the data were qualitatively consistent with predictions of a normalization model. Here we varied input contrasts over a wide range to obtain a more complete profile of masking responses that we used to test different parameterizations of the divisive normalization model.

Materials and Methods

Observers. Ten neurotypical observers (4 female) with normal or corrected-to-normal visual acuity participated. A local ethics review board approved the recruitment and experiment procedures before the start of the project.

Display and stimuli. Stimuli were displayed on a 19" monitor (Electron Blue; LaCie) at a spatial resolution of 800×600 pixels, 72 Hz vertical refresh rate, and mean luminance of 81 cd/m^2 . The nonlinear voltage versus luminance response of the monitor was corrected in software. Stimuli were generated and presented using an in-house display system (PowerDIVA) with high temporal precision. Subjects viewed the monitor from a distance of 63 cm, giving a view angle of $32.3^\circ \times 24.2^\circ$ ($2.4'$ per pixel).

Stimuli consisted of one or two superimposed random checkerboard patterns, with each check measuring $12 \times 12'$ (Fig. 1). To minimize EEG signal cancellation and to thus improve the accuracy of source localization, we presented the pattern in the lower right quadrant of the display. The same checkerboard pattern was used throughout the experiment for all subjects. In some of the trials, a second mask pattern was also present. The mask pattern was of the same size as the original (test), generated independently and superimposed on the test. The contrasts of the test and mask were modulated sinusoidally from zero to some positive value c , thus around a mean contrast of $c/2$. For the test, c was stepped from 0.5% to 47% in 10 equal logarithmic intervals in each trial. For the mask, c was fixed at one of four values (0%, 5%, 10%, 20%) in each trial. The frequencies of modulation were 5.14 Hz and 7.2 Hz for test and mask, respectively. The interval between steps was 0.97 s (the smallest common multiple of the period of the two input frequencies) for a total trial

duration of 9.7 s. Twenty repeats at each mask contrast (total of 80 trials) were randomized.

Each trial began with a central fixation mark, followed by presentation of the mask and test stimuli for 9.7 s. Observers' attention was controlled using a stream of letters shown at the center of the display, among which a target letter was to be detected from distractors.

Steady-state VEPs recording and preprocessing. EEG signals were recorded using 128-channel HydroCell Sensor Nets (Electrical Geodesics). Signals were recorded with a vertex physical reference, amplified with a gain of 1000, bandpass filtered between 0.1 and 50 Hz, and digitized at a sampling rate of 432 Hz. At the end of each experimental session, the 3D locations of each sensor and of three fiducials (nasion, left and right preauricular) were digitized using a Fastrack 3D digitizer (Polhemus).

Artifact rejection was processed off-line in two stages. In the first stage, raw data were evaluated sample by sample to determine those that exceeded a threshold ($\sim 30 \mu\text{V}$). Noisy channels that had $>10\%$ of the samples exceeding the threshold were replaced by the average of the six nearest neighbors. In a second stage, individual channels were evaluated sample by sample, and epochs that contained large number of sensors (>7) exceeding a threshold ($\sim 60 \mu\text{V}$) were rejected. Typically, rejected data corresponded to periods of eye movements or blinks. After artifact rejection, the EEG was re-referenced to the common average of all the sensors.

Scalp EEG activity was converted to cortical current density using a method of EEG source reconstruction described in detail previously (Cottareau et al., 2011). In brief, the method begins with boundary element model of tissues in the head constructed from each subject's MRI scans. Visual areas were defined by a separate procedure based on retinotopic mapping using fMRI (Engel et al., 1997). Cortical activity in early visual cortex (V1) was calculated using a L2 minimum norm solution with sources constrained to the location and orientation of the cortex. Additional constraints on source localization included the restriction to the dorsal parts of the hemisphere contralateral to the stimulus and a weighting scheme in which visual areas received twice the weight of nonvisual areas.

Response waveforms in V1 were converted to the frequency domain via a discrete Fourier transform with a resolution of 1.03 Hz. When pooling across subjects, the responses were averaged coherently (i.e., taking into account both amplitude and phase).

Contrast response modeling. The response to the sweep of contrast comprised a contrast response function (CRF). We extend a well established description of the CRF—the hyperbolic ratio function (Naka and Rushton, 1966; Albrecht and Hamilton, 1982)—to account for multiple frequency components in our steady-state paradigm. This model describes the response of a neuron or a population having an accelerating response nonlinearity whose input can be modulated by a divisive component arising from the combined responses of all other neighboring neurons (a gain pool). Previous normalization models of this type (Albrecht and Hamilton, 1982; Heeger, 1992; Carandini et al., 1997; Busse et al., 2009) typically operate on a scalar input representing the stimulus contrast and produce a scalar output representing the response amplitude, as shown in Equation 1:

$$R = R_m \frac{c^n}{c^n + \sigma^n}, \quad (1)$$

where R is the response, c is the stimulus contrast, and R_m , n , and σ are parameters representing the response maximum and the minimum, the exponent, and the contrast producing the half-maximal response, respectively. We extend this model to incorporate temporal dynamics in two stages. First, following Candy et al. (2001), we let the input be a time series, representing the temporal modulation of stimulus contrast. The output is also a function of time:

$$R(t) = R_m \frac{[c(t)]^n}{[c(t)]^n + \sigma^n}, \quad (2)$$

where $c(t)$ is the temporal modulation of the stimulus contrast—in this case, represented by the sum of the two input sinusoids. Notice that the gain pool, represented in the denominator of Equation 2, is time-varying,

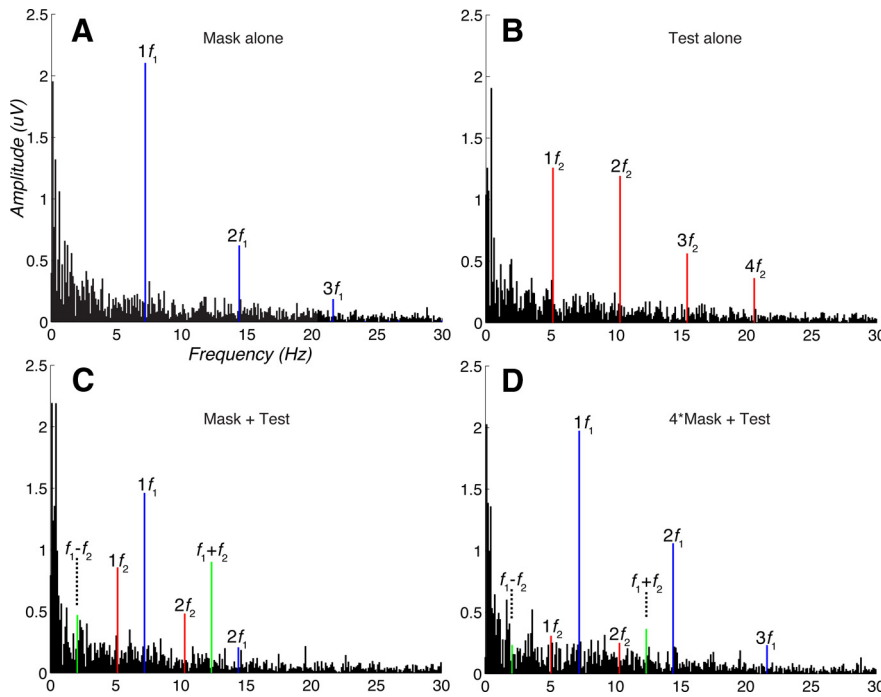


Figure 2. Amplitude spectrum of the SSVEPs recorded at Oz. The labels denote the frequency of the response components in terms of multiples of the input frequencies. **A, B**, Mask (f_1) and test (f_2) stimulus presented separately at 10% contrast. The stimulus-driven components are clearly greater than the background EEG and are seen at integer multiples of the input frequencies f_1 (blue) and f_2 (red). In these panels, the resolution of the spectrum is 0.103 Hz, 10 times better than the resolution available in the main experiment, because the analysis window is 10 times as long. Note the precise isolation of the SSVEPs to specific frequencies in the spectrum. **C**, Concurrent presentation of mask and test of equal contrast (10%). Responses corresponding to harmonics of the stimulus frequencies are present, in addition to intermodulation terms (green), some of which are labeled. **D**, Concurrent presentation of the mask and test, with the mask contrast at 40% and the test contrast at 10%. The pattern of the spectrum closely resembles the responses elicited by the mask stimulus alone (cf. **A**). The frequencies corresponding to the test stimulus are notably absent even though the test stimulus is shown at the same contrast in both **B** and **D**.

as neurons in the gain pool receive the same input as the neuron. In addition, we allow the exponents of the excitation and inhibition components, p and q , respectively, to be fitted separately (Foley, 1994; Chen et al., 2001; Xing and Heeger, 2001; Peirce, 2007). To produce good fits to the CRF at the high contrasts, it is necessary to allow σ to vary with mask contrast (Ross and Speed, 1991).

This model may be compared with another variant of normalization in which the denominator term, the gain pool, is time independent (Bonin et al., 2005) and reflects a spatiotemporal integration of local stimulus contrast, c_{local} . This model is described by the following:

$$R(t) = R_m \frac{[c(t)]^p}{(c_{\text{local}})^q + \sigma^q}, \quad (3)$$

where $c_{\text{local}} \propto \sqrt{c_{\text{mask}}^2 + c_{\text{test}}^2}$, and c_{mask} and c_{test} are the contrasts of the mask and test, respectively (Carandini, 2004; Bonin et al., 2006).

The second modification in our model describes the temporal dynamics of the gain pool response. In the Candy-style model (Eq. 2), there is no temporal integration so that the response of the gain pool contains the full temporal spectrum. In comparison, the Bonin-style model (Eq. 3) integrates over space and time so that the gain pool response is a constant. These are two ends of a continuum. To generalize, we allowed the temporal integration window of the gain pool to be a free parameter. This model is described by the following:

$$R(t) = R_m \frac{[c(t)]^p}{f(t) + \sigma^q}, \quad (4)$$

where $f(t)$ is a temporally filtered version of the gain pool response:

$$f(t) = [c(t)]^q * h(t). \quad (5)$$

The filter impulse response is assumed to be a decaying exponential, $h(t) = \frac{1}{\tau} e^{-t/\tau}$. This represents a low-pass filter with a half-max cut-off frequency of $1/\tau$.

Model fitting was done by a numerical search (MATLAB function lsqnonlin) to minimize the quantity:

$$\chi^2 = \sum_i \frac{(R_i - \hat{R}_i)^2}{s^2} \quad (6)$$

where R_i is the response amplitude at the i th combination of contrasts and frequency, \hat{R}_i is the model prediction, and s is the standard error of the response.

Results

Responses to periodic stimuli are conveniently described by their frequency spectrum. An amplitude spectrum of the steady-state VEPs (SSVEPs) recorded from an example subject is shown in Figure 2. Here, to simplify the illustration, both test and mask had a fixed contrast modulation (i.e., no contrast sweep) and a single channel centered on the occipital cortex (Oz) is depicted. In response to the test or mask alone (Fig. 2*A, B*), the stimulus-driven response was precisely identified in the spectrum, shown as large peaks of activity at integer multiples of the input frequencies (self terms; Fig. 2, red and blue lines). When mask and test of equal contrast were presented concurrently (Fig. 2*C*), self terms can be seen, as well as additional stimulus-driven components

at frequencies equal to low-order sums and differences of the stimulus frequencies [intermodulation (IM) terms; Fig. 2*C*, green lines]. Finally, with different mask and test contrasts (e.g., mask four times the test contrast; Fig. 2*D*), the response closely resembled that elicited by the stronger stimulus alone. Remarkably, although the test stimulus in both Figure 2*B* and Figure 2*D* had the same contrast, its self terms components (nf_2) were notably absent in the presence of a stronger mask stimulus. Moreover, intermodulation terms were reduced to the noise level. Hence, the winner-take-all behavior is clearly manifest in multiple spectral components of the SSVEP responses.

Profiles of masking in frequency domain

We focused our analysis on signals from early visual cortex (V1) because contrast representation is most thoroughly studied at this level (Albrecht and Hamilton, 1982; Carandini et al., 2005). To characterize the possible interactions between the test and mask stimuli, we sampled combinations of test and mask over a large range of relative contrast values and looked at the responses corresponding to self and intermodulation terms (Regan and Regan, 1988).

We first examined the self terms. The first harmonic responses of all subjects were averaged coherently and the amplitude of the mean response was plotted against the test stimulus contrast (Fig. 3*A*). Each of the CRFs in Figure 3 corresponds to a fixed mask contrast. With increasing mask strength, the CRF was shifted rightwards, consistent with a reduction in the effective contrast of

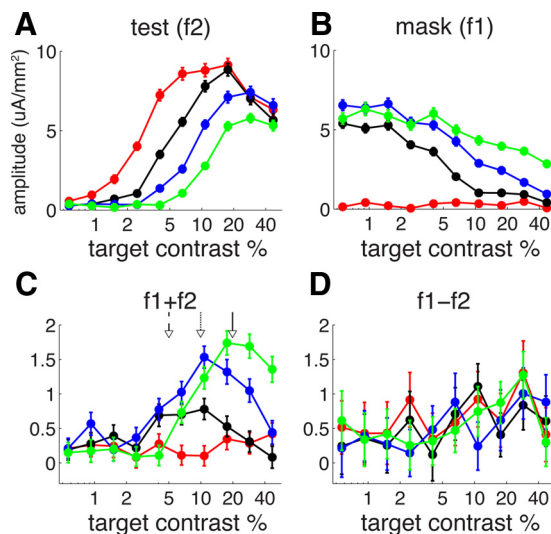


Figure 3. Group mean ($n = 10$) of cortical current amplitude in V1. Error bars denote SEM. Colors denote mask contrast: red, 0%; black, 5%; blue, 10%; green, 20%. **A**, Response to the test stimulus (measured at the frequency f_2). **B**, Response to the mask stimulus (measured at the frequency f_1). **C**, Amplitude of the second-order sum IM term (measured at the frequency $f_1 + f_2$). Arrows indicate the point of equality between test and mask contrasts. **D**, Amplitude of the second-order difference term (measured at the frequency $f_1 - f_2$).

the test. In turn, the mask response was reduced by the test stimulus (Fig. 3B)—in the presence of increasing test contrast, the mask response decreased monotonically. These patterns are consistent with previous studies of masking (Freeman et al., 2002; Busse et al., 2009).

Next, we examined the second order ($f_1 + f_2$) component because it was the dominant IM response in our data (Fig. 2C). In the same manner as the self terms, the magnitude of this component in response to combinations of test and mask contrast is shown in Figure 3C. Not surprisingly, IM response was absent when the test was presented alone (Fig. 3, red curve). When both test and mask were present, the response amplitude as a function of test contrast was nonmonotonic: it increased with test contrast but peaked and decreased thereafter to baseline. When test contrast was much greater than the mask contrast, IM response was negligible, as though only one stimulus was present—a winner-take-all situation. The peak of the response occurred when the test and mask contrasts were equal (Fig. 3C, arrows). Hence, an additional signature of WTA is the absence of IM components. This dissociation is revealing since both WTA and the generation of IM responses might have depended on the same nonlinear mechanism.

Finally, the response at the difference intermodulation frequency ($f_1 - f_2$) is shown in Figure 3D. The amplitude of the response at this frequency was smaller than that at the sum frequency. In Figure 3, C and D, the response with mask contrast of zero (red curves) can be taken as a measure of the noise level because no IM response is expected. Although the noise level was higher at the difference than the sum frequency, a signal of comparable magnitude to that measured at the sum frequency would have been readily detectable. In fact, none of the mask contrasts elicited a response that rose above the noise level (Fig. 3D). This proved to be an important constraint on our models, as we demonstrate below.

In summary, we find that components of the spectrum carry distinct signature of visual masking. In the next section, we address whether the normalization process can account for these results.

We should point out that the drop-off to zero of the IM term is qualitatively different from the small roll-off of test self terms at high contrasts (Fig. 3A). The latter is known as response supersaturation, reported in some visual neurons (Albrecht and Hamilton, 1982; Li and Creutzfeldt, 1984; Peirce, 2007) and in human VEP (Tyler and Apkarian, 1985; Burr and Morrone, 1987). We note in passing that supersaturation is not attributable to a slow adaptation process (Carandini and Ferster, 1997) because supersaturation is present at the very beginning of each trial (data not shown).

Dynamics of the gain pool

Previous instantiations of normalization models could not predict the full range of results presented above, in part because of limitations in the intended scope of these models with respect to either temporal dynamics or input contrast. Here, we develop a variant of the normalization model that explains the full range of frequency-domain responses.

Following Candy et al. (2001), we constructed a model in which the input corresponded to the temporal waveforms of the contrast modulation. As the output of the model only depends on the instantaneous input, we call this the memory-less model. We fit this model to the first-order self terms (f_1, f_2) and the second-order sum IM term ($f_1 + f_2$) for all mask contrasts simultaneously (Fig. 4A; 120 data points; for model parameter values, see Table 1) because these terms contained most of the response signals. The predictions (Fig. 4A, lines) accounted for 94% of the variance in the data and captured the qualitative features of the masking behavior in the self terms (Fig. 4A, first and second columns); that is, the curves shifted laterally with mask contrast. However, the predictions depart from the data with respect to the sum IM term, specifically because the peak of the response did not fall on the point of equality (Fig. 4A, third column).

In an alternative normalization model (Bonin et al., 2005), the gain pool response is constructed from spatiotemporal integration of the stimulus over a suppressive field. Contrary to the memory-less model, this one has effectively long-memory. The results of fitting this model to our data are shown in Figure 4B. Again, the model captured the shifting of the self term responses as mask contrast increased (93% of the variance explained). However, for the sum IM term, the model predicted responses that were much smaller than observed.

The Candy and Bonin models differ in the extent of temporal integration in the gain pool signal. In the Candy model, this integration window is infinitesimally short or absent, while in the Bonin model, the window is effectively very long. To assess the effect of the duration of integration, we fitted it as an additional model parameter (see Materials and Methods, above). The best fitting parameters had a time constant of 26 ms, and the model (short-memory) produced excellent fit to the data, accounting for 96% of the variance (Fig. 4C).

Following Cavanaugh et al. (2002), we characterized the goodness-of-fit of these three models using a normalized χ^2 measure, which takes into account the number of model parameters:

$$\chi_N^2 = \frac{\chi^2}{df} \quad (7)$$

where χ^2 is given by Equation 6, df is the number of degrees of freedom in the model. The best fitting model is the one with the lowest χ_N^2 . The normalized χ^2 for the memory-less, long-memory, and the short-memory models are 2.26, 2.51, and 1.47, respectively. All three models show similar performance in pre-

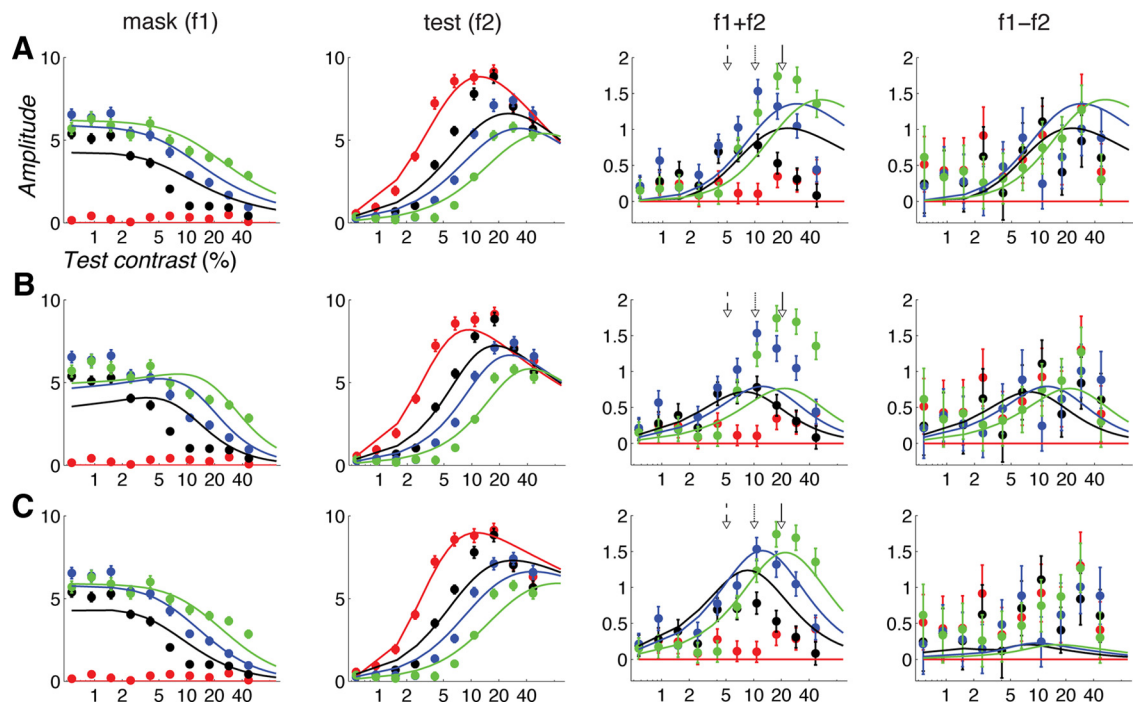


Figure 4. *A–C*, Fitting of three variants of the normalization model (*A*, memory-less; *B*, long-memory; *C*, short-memory) to the contrast response data from Figure 3. Colors correspond to different mask contrast: red, no mask; black, 5%; blue, 10%; green, 20%. All three models show similar goodness-of-fit in the mask (first column) and test (second column) terms. However, the short-memory model best captures the characteristics of the intermodulation terms, including the peak of the sum response function when the test and mask contrasts are equal (third column) and the diminished signal/noise at the difference frequency relative to the sum (fourth column).

Table 1. Values of fitted parameters

	R_{\max}	σ_{mask_1} (%)	σ_{mask_2}	σ_{mask_3}	σ_{mask_4}	p	q	τ (ms)
Memory-less	0.05 (0.02, 0.07)	3.5 (1.7, 5.3)	7.1 (4.4, 9.7)	9.2 (5.9, 12.4)	17 (12, 21)	1.49 (1.2, 1.8)	1.52 (1.2, 1.8)	n/a
Long-memory	0.08 (0.03, 0.12)	3.9 (2.3, 5.4)	8.2 (6.1, 10)	11 (7.9, 13)	15 (11, 19)	1.6 (1.4, 1.8)	1.9 (1.7, 2.1)	n/a
Short-memory	23 (12, 34)	37 (8, 66)	66 (23, 111)	81 (30, 133)	128 (51, 204)	2.2 (1.7, 2.7)	2.4 (1.9, 2.9)	26 (16, 36)

Values in parentheses are 95% confidence intervals.

dicting the self terms; however, the short-memory model best captures the characteristics of the intermodulation term. For the sum ($f_1 + f_2$) term, the short-memory model correctly identifies the shift of the peak of the function with mask contrast. Furthermore, only this model predicts an asymmetry between the sum and difference IM responses (Fig. 4, right). These results demonstrate that inclusion of the intermodulation terms is critical to constraining the model.

Discussion

Few previous studies have used the frequency-tagging technique to study masking (Burr and Morrone, 1987; Bonds, 1989; Ross and Speed, 1991; Candy et al., 2001; Bonin et al., 2005; Busse et al., 2009). Only Candy et al. (2001) have studied the intermodulation term that we found to be most effective in discriminating among models. The other studies focused on the self terms. We have shown that the combination of spectral components of a neural population response carry distinctive signatures of contrast masking and gain control. Specifically, second-order IM terms in SSVEP is maximal when two inputs are of the same contrast and negligible when they are markedly different. In comparison, the self terms have a different profile—the response increases through the point of equality. In addition, the IM terms are critical to constraining a model for describing masking and gain control. Our results are well described by a divisive normal-

ization model that includes additional temporal integration in the generation of the gain pool.

The findings presented here do not depend critically on EEG source imaging. Data from a single electrode, Oz, referenced to the average are very similar to V1. This is not surprising since we have focused on contrast masking, an early visual system process, and Oz largely reflects the activity of cells in V1 (Ales et al., 2010). However, in other paradigms it may be desirable to examine additional visual areas and our technique can be extended to this situation easily. We also note that in some people, the occipital lobe extends ventrally in a nonstereotypical manner and electrode Oz will then sit over more dorsal visual areas, such as V3a. Our technique eliminated potential errors of this type.

Dynamics of the gain pool

The temporal dynamics of the gain pool are strongly constrained by the observed data. Modeling shows that a temporal integration stage with a time constant of ~ 26 ms must be applied to the normalization signal to fit the data. This suggests that normalization is rapid, but not instantaneous. There are some lines of support in the literature. First, there is evidence for temporal integration in retinal gain control mechanisms (Shapley and Victor, 1981; Victor, 1987), and second, the proposed temporal integration is compatible with the delayed onset of suppression observed in the cortex (Bair et al., 2003; Smith et al., 2006). In the

retina, Victor (1987) proposed an elegant model of gain control comprised of a series of low-pass filters, an adaptive high-pass filter, and spike transduction. The high-pass filter was characterized by a time constant, which in turn depended on a neural measure of stimulus contrast. Victor (1987) showed that a strictly linear model and a quasilinear (with long integration time) model were both inadequate. Instead, the best-fitting model required a time constant in the range of 5 to 25 ms.

The variants of normalization models we discussed in this paper make predictions for the onset of suppression. An infinitesimally short integration period (Candy et al., 2001) predicts that the gain pool signal is fully developed at the outset, resulting in no delay of the suppression effect relative to the response onset. Conversely, a finite integration constant predicts a delay in the onset of suppression. In macaque V1, neurons show a delay in response suppression relative to the earliest response to a visual stimulus (Bair et al., 2003; Smith et al., 2006). There are two qualitatively different types of suppressive interactions: overlay and surround suppression (Petrov et al., 2005). The delay in overlay suppression was 13 ms on average relative to response offset (Smith et al., 2006). Surround suppression was slower than overlay suppression by an additional 12 ms (Smith et al., 2006). The large spatial extent of our stimulus likely engaged both suppression mechanisms, and the estimate of temporal integration is consistent with the delay of suppression measured in single cells.

Where might be the neural loci for this dynamic gain control? Studies using pairs of oriented gratings revealed that the second-order intermodulation response depends on the relative orientation of the stimulus (Regan and Regan, 1987; Candy et al., 2001; Baker et al., 2011). This implies that an orientation-sensitive mechanism is involved in the generation of the IM response. Furthermore, intermodulation response can be obtained from component gratings presented dichoptically, implying that IM response are generated after inputs from the two eyes have been combined (Brown et al., 1999; Norcia et al., 2000). However, these results do not rule out a precortical contribution to shaping the IM response. In particular, a recent model of binocular interaction posits that excitatory and inhibitory interactions between eyes take place in cortical and precortical channels, respectively (Zhang et al., 2011). If IM response reflects the excitatory interaction, this model could account for the dissociation between masking and IM response in our data, i.e., the absence of IM when masking was strong (Fig. 3), because strong inhibition of one input by the other at a precortical site precludes downstream excitatory interactions.

One limitation of our model is its relative deficiency in predicting the phase of SSVEP responses. While the short-memory model predicts some aspects of the measured data, such as a decrease in phase-lag with increasing contrast that the memoryless and long-memory models do not, it does not correctly predict absolute phase correctly. Several factors complicate the interpretation of phase in our dataset. Unlike amplitude, phase measurement of weak signals yields unpredictable values and so the phase of responses to low-contrast inputs is poorly defined. Furthermore, intersubject variability can produce indeterminate confidence bounds (i.e., $>2\pi$) because of the circular nature of phase data. Finally, response phase is dependent on conduction delays in the visual pathways, which we did not include in the

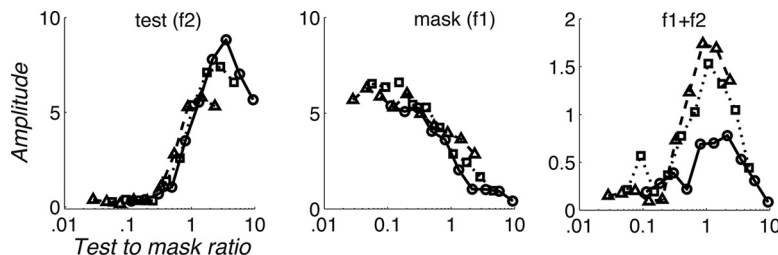


Figure 5. Data from Figure 3 replotted as a function of the ratio of the test to mask contrast. Left and middle, The contrast response functions essentially overlap for different mask contrasts. Right, Peak of the IM response occurs at a ratio of one.

model. A more detailed model of membrane conductances, such as shunting inhibition (Carandini and Heeger, 1994; Carandini et al., 1997; Sit et al., 2009) may be a way to augment the model. Nonetheless, the modeling here does demonstrate that it is necessary to consider the dynamics of the gain control pool to account for the interactions underlying masking.

Winner-take-all, invariance, and normalization

The normalization model exhibits a variety of behaviors based on the relative input contrasts. On the one hand, when the test and mask contrast are different, a winner-take-all operation is apparent (Fig. 2). On the other hand, when their contrasts are similar, gain control is apparent from the shift of the contrast response functions (Fig. 3). The purpose of gain control may be to adjust the sensitivity such that the response remains invariant with respect to changes in the environment.

In support of this, we identify a contrast–contrast invariance (Fig. 5), which appears to be a strong form of contrast normalization as traditionally studied with psychophysical techniques. In the psychophysical literature, the appearance of a test patch depends on the relative contrast of the patch and its surround (Ejima and Takahashi, 1985; Chubb et al., 1989; Cannon and Fullenkamp, 1991; Xing and Heeger, 2000). Relative contrast is known to affect the sensitivity of stereo depth, motion, and vernier perception in that maximal sensitivity is achieved when the contrast ratio between corresponding images is one (Halpern and Blake, 1988; Stevenson and Cormack, 2000). In Figure 5, we show that neural responses depend only on the ratio of the test and mask contrasts over a very wide range of absolute contrasts.

Most strikingly, sensitivity to the test (Fig. 5, left) is precisely centered on a contrast ratio of 1 so that it is positioned to give the maximum differential response. This contrast–contrast invariance was also evident in the second-order IM response. The magnitude of the IM term is centered so that at a ratio of 1, the two signals are maximally mixed and, as the ratio deviates from unity, WTA behavior emerges (Fig. 5, right). This behavior may provide a means for selection between competing responses. Together, these data show that at a population level, neurons in visual cortex operate on a representation of relative rather than absolute contrast and that this invariance can be understood in the framework of normalization provided one takes into account its dynamics.

References

- Albrecht DG, Hamilton DB (1982) Striate cortex of monkey and cat: contrast response function. *J Neurophysiol* 48:217–237.
- Ales JM, Yates JL, Norcia AM (2010) V1 is not uniquely identified by polarity reversals of responses to upper and lower visual field stimuli. *Neuroimage* 52:1401–1409.
- Anderson JS, Lampl I, Gillespie DC, Ferster D (2000) The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science* 290:1968–1972.

- Bair W, Cavanaugh JR, Movshon JA (2003) Time course and time-distance relationships for surround suppression in macaque V1 neurons. *J Neurosci* 23:7690–7701.
- Baker TJ, Norcia AM, Candy TR (2011) Orientation tuning in the visual cortex of 3-month-old human infants. *Vision research* 51:470–478.
- Bonds AB (1989) Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Vis Neurosci* 2:41–55.
- Bonin V, Mante V, Carandini M (2005) The suppressive field of neurons in lateral geniculate nucleus. *J Neurosci* 25:10844–10856.
- Bonin V, Mante V, Carandini M (2006) The statistical computation underlying contrast gain control. *J Neurosci* 26:6346–6353.
- Boynton GM (2009) A framework for describing the effects of attention on visual responses. *Vision Res* 49:1129–1143.
- Brown RJ, Candy TR, Norcia AM (1999) Development of rivalry and dichoptic masking in human infants. *Investigative ophthalmology and visual science* 40:3324–3333.
- Burr DC, Morrone MC (1987) Inhibitory interactions in the human vision system revealed in pattern-evoked potentials. *J Physiol* 389:1–21.
- Busse L, Wade AR, Carandini M (2009) Representation of concurrent stimuli by population activity in visual cortex. *Neuron* 64:931–942.
- Candy TR, Skoczenski AM, Norcia AM (2001) Normalization models applied to orientation masking in the human infant. *J Neurosci* 21:4530–4541.
- Cannon MW, Fullenkamp SC (1991) Spatial interactions in apparent contrast: inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations. *Vision Res* 31:1985–1998.
- Carandini M (2004) Receptive fields and suppressive fields in the early visual system. In: *The cognitive neurosciences*, 3rd Edition (Gazzaniga MS, ed), pp 313–326. Cambridge, MA: MIT.
- Carandini M (2007) Melting the iceberg: contrast invariance in visual cortex. *Neuron* 54:11–13.
- Carandini M, Ferster D (1997) A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. *Science* 276:949–952.
- Carandini M, Heeger DJ (1994) Summation and division by neurons in primate visual cortex. *Science* 264:1333–1336.
- Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. *Nat Rev Neurosci* 13:51–62.
- Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* 17:8621–8644.
- Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC (2005) Do we know what the early visual system does? *J Neurosci* 25:10577–10597.
- Cavanaugh JR, Bair W, Movshon JA (2002) Nat and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J Neurophysiol* 88:2530–2546.
- Chen CC, Kasamatsu T, Polat U, Norcia AM (2001) Contrast response characteristics of long-range lateral interactions in cat striate cortex. *Neuroreport* 12:655–661.
- Chubb C, Sperling G, Solomon JA (1989) Texture interactions determine perceived contrast. *Proc Natl Acad Sci U S A* 86:9631–9635.
- Cottareau BR, McKee SP, Ales JM, Norcia AM (2011) Disparity-tuned population responses from human visual cortex. *J Neurosci* 31:954–965.
- Creutzfeldt OD (1977) Generality of the functional structure of the neocortex. *Naturwissenschaften* 64:507–517.
- DeAngelis GC, Robson JG, Ohzawa I, Freeman RD (1992) Organization of suppression in receptive fields of neurons in cat visual cortex. *J Neurophysiol* 68:144–163.
- Douglas RJ, Martin KA (2004) Neuronal circuits of the neocortex. *Annu Rev Neurosci* 27:419–451.
- Ejima Y, Takahashi S (1985) Apparent contrast of a sinusoidal grating in the simultaneous presence of peripheral gratings. *Vision Res* 25:1223–1232.
- Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex* 7:181–192.
- Foley JM (1994) Human luminance pattern-vision mechanisms: masking experiments require a new model. *J Opt Soc Am A Opt Image Sci Vis* 11:1710–1719.
- Freeman TC, Durand S, Kiper DC, Carandini M (2002) Suppression without inhibition in visual cortex. *Neuron* 35:759–771.
- Greenlee MW, Heitger F (1988) The functional role of contrast adaptation. *Vision Res* 28:791–797.
- Halpern DL, Blake RR (1988) How contrast affects stereoacuity. *Perception* 17:483–495.
- Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9:181–197.
- Kouh M, Poggio T (2008) A canonical neural circuit for cortical nonlinear operations. *Neural Comput* 20:1427–1451.
- Legge GE, Foley JM (1980) Contrast masking in human vision. *J Opt Soc Am* 70:1458–1471.
- Li CY, Creutzfeldt O (1984) The representation of contrast and other stimulus parameters by single neurons in area 17 of the cat. *Pflügers Arch* 401:304–314.
- Miller KD, Troyer TW (2002) Neural noise can explain expansive, power-law nonlinearities in neural response functions. *J Neurophysiol* 87:653–659.
- Morrone MC, Burr DC, Maffei L (1982) Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proc R Soc Lond B Biol Sci* 216:335–354.
- Naka KI, Rushton WA (1966) S-potentials from colour units in the retina of fish (*Cyprinidae*). *J Physiol* 185:536–555.
- Norcia AM, Harrad RA, Brown RJ (2000) Changes in cortical activity during suppression in stereoblindness. *Neuroreport* 11:1007–1012.
- Peirce JW (2007) The potential importance of saturating and supersaturating contrast response functions in visual cortex. *J Vis* 7:13.
- Petrov Y, Carandini M, McKee S (2005) Two distinct mechanisms of suppression in human vision. *J Neurosci* 25:8704–8707.
- Regan D, Regan MP (1987) Nonlinearity in human visual responses to two-dimensional patterns, and a limitation of Fourier methods. *Vision Res* 27:2181–2183.
- Regan MP, Regan D (1988) A frequency domain technique for characterizing nonlinearities in biological systems. *J Theor Biol* 133:293–317.
- Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185.
- Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019–1025.
- Ross J, Speed HD (1991) Contrast adaptation and contrast masking in human vision. *Proc Biol Sci* 246:61–69.
- Schwartz O, Simoncelli EP (2001) Natural signal statistics and sensory gain control. *Nat Neurosci* 4:819–825.
- Shapley RM, Victor JD (1981) How the contrast gain control modifies the frequency responses of cat retinal ganglion cells. *J Physiol* 318:161–179.
- Sit YF, Chen Y, Geisler WS, Mäkeläinen R, Seidemann E (2009) Complex dynamics of V1 population responses explained by a simple gain-control model. *Neuron* 64:943–956.
- Smith MA, Bair W, Movshon JA (2006) Dynamics of suppression in macaque primary visual cortex. *J Neurosci* 26:4826–4834.
- Stevenson SB, Cormack LK (2000) A contrast paradox in stereopsis, motion detection, and vernier acuity. *Vision Res* 40:2881–2884.
- Tyler CW, Apkarian PA (1985) Effects of contrast, orientation and binocularity in the pattern evoked potential. *Vision Res* 25:755–766.
- Victor J, Shapley R (1980) A method of nonlinear analysis in the frequency domain. *Biophys J* 29:459–483.
- Victor JD (1987) The dynamics of the cat retinal X cell centre. *J Physiol* 386:219–246.
- Xing J, Heeger DJ (2000) Center-surround interactions in foveal and peripheral vision. *Vision Res* 40:3065–3072.
- Xing J, Heeger DJ (2001) Measurement and modeling of center-surround suppression and enhancement. *Vision Res* 41:571–583.
- Zhang P, Bobier W, Thompson B, Hess RF (2011) Binocular balance in normal vision and its modulation by mean luminance. *Optom Vis Sci* 88:1072–1079.